Airspace Research and Development Portfolio Assessment of Urban Air Mobility using Knowledge Graph Data Science

Nipa Phojanamongkolkij[1]
NASA Langley Research Center, Hampton, VA, 23681, USA

Joshua Bonner[2]
Science Applications International Corporation, Hampton, VA, 23681, USA

Ian Levitt[3]
NASA Langley Research Center, Hampton, VA, 23681, USA

National Aeronautics and Space Administration (NASA) is spearheading an innovative digital engineering approach to integrate, communicate, and facilitate the research of Urban Air Mobility (UAM) operations. The UAM vision is one in which advanced technologies and new operational procedures enable practical and cost-effective air transport as an integrated mode of movement of people and goods throughout metropolitan areas. To safely support UAM operations at scale in the National Airspace System (NAS), NASA's Air Traffic Management-Exploration (ATM-X) project has been conducting research that evolves the UAM air traffic management system towards a highly automated and operationally flexible system of the future. The complexity of UAM airspace evolution to accommodate the increasing tempo of UAM operations over time is managed through the UAM airspace research roadmap, which is a system engineering approach to the R&D of complex system-of-systems, where system's interdependencies make it nearly impossible to define requirements for individual elements of the system in isolation. These interdependencies form a knowledge graph (node-link network) with a highly complex structure far beyond the human user's ability to extract insights for project management's research portfolio assessment. This study applies advanced data analytics in knowledge graph to the UAM knowledge graph to facilitate the portfolio assessment.

I. Introduction

Advanced Air Mobility (AAM) encompasses a range of innovative operational and technological changes to aviation (electric aircraft, increasingly automated aircraft, increasingly automated airspace operations, etc.) that transform aviation's role in the everyday movement of people and goods. There are multiple associated concepts and use cases for AAM, all interrelated, including small Unmanned Aircraft System (UAS) Traffic Management (UTM), Upper-Class E Traffic Management (ETM), Extensible Traffic Management (xTM), Regional Air Mobility (RAM), and Urban Air Mobility (UAM). These AAM operations must integrate with traditional Air Traffic Management (ATM) operations and non-aviation modes of transportation and logistics.

National Aeronautics and Space Administration (NASA) Langley Research Center (LaRC) is spearheading an innovative digital engineering approach to integrate, communicate, and facilitate the research of AAM operations. The Knowledge-based Digital Platform (KbDP) is a concept being developed that ties the workflows of Project

---

[1] Senior Systems Engineer, Engineering Integration, Engineering Directorate.

[2] Data Scientist, Data Science Services, Information Data, and Analytics Services.

[3] Principal Engineer, Crew Systems and Aviation Operations, Research Directorate.

Managers (PM), Principal Investigators (PI), and System Engineers (SE) together across organizational boundaries. Imagine the knowledge platform allowing any user to access the latest AAM research and development (R&D) architecture, concepts, and requirements, including supporting technical reports and standards from which the R&D contents were derived. At the heart of the KbDP is an information database defined by mathematical, data science, and system engineering principles. The overarching vision for the KbDP Concept for AAM R&D is a substantial undertaking. The initial concept and implementation will focus on UAM operations to tractably learn and adjust the concept with a manageable database. Lessons learned and best practices with these UAM operations will enable successful scalability to AAM R&D or even to the entire modes of transportation and logistics.

The UAM operations cover a subset of the AAM concepts, namely those that use electric Vertical Takeoff and Landing (eVTOL) aircraft to provide air-taxi or package delivery services to the public over densely populated cities and the urban periphery, including flying between local, regional, intraregional, and urban locations. The UAM vision is one in which advanced technologies and new operational procedures enable practical and cost-effective air transport as an integrated mode of movement of people and goods throughout metropolitan areas. To safely support UAM operations at scale in the National Airspace System (NAS), NASA's Air Traffic Management-Exploration (ATM-X) project has been conducting research that evolves the UAM air traffic management system towards a highly automated and operationally flexible system of the future.

The complexity of UAM airspace evolution to accommodate the increasing tempo of UAM operations over time requires a plan to effectively organize, integrate, and communicate NASA's research and development. The UAM airspace research roadmap (Ref. [1]), or herein roadmap, is a system engineering approach to the R&D of complex system-of-systems, where interdependencies make it nearly impossible to define requirements for individual elements of the system in isolation. These interdependencies form a knowledge graph (node-link network) with a highly complex structure (partial network displayed in Figure 1) far beyond the subject-matter-expert's ability to extract insights in order to answer project management's questions about critical airspace requirements. Does the ATM-X project's portfolio address any of these requirements? What level of maturity are these requirements? On the other hand, there is more advanced data analytics in the knowledge graph that should be applied to the current UAM knowledge graph to address these questions.

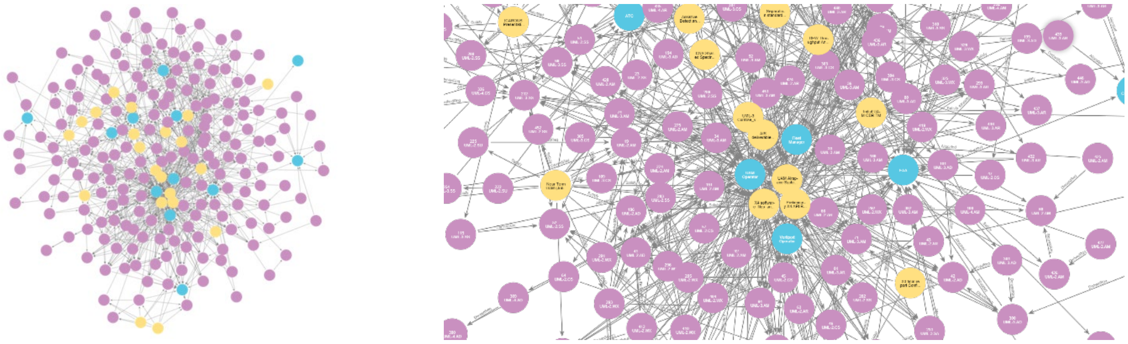


Fig. 1 Partial UAM Knowledge Graph Structure relevant to this study where purple nodes are requirements, blue nodes are system actors, and yellow nodes are project's deliverables (the left figure for the bird-eye view and the right for the zoom-in view)

The organization of the paper is as follows. Section II describes the methodology used to address aforementioned questions. The results are given in Section III and, lastly, Section IV provides a conclusion and the next steps.

II. Portfolio Analysis Methodology

A. Knowledge Graph Ontology Definition

The knowledge graph structure in Figure 1 follows the ontology definition given in Figure 2. The ATM-X project has established programmatic milestone tasks in the integrated master schedule (IMS) using MS Project software. These milestones are reported to the project's stakeholders as a mean to deliver the project's progress toward stakeholder's expectations. Whenever the researchers complete the milestone task, its deliverables (e.g., software, user manual, technical report) are reviewed by a systems engineer (SE). The SE qualitatively assesses which of the roadmap requirements that the deliverables ***satisfy*** and qualitatively evaluate the Technology Readiness Level (TRL[4]) of the deliverables using the well-established criteria given in Ref. [3]. One deliverable can satisfy multiple roadmap requirements, and similarly one requirement can be satisfied by multiple deliverables, resulting in many-to-many ***satisfy*** relationship as depicted in the figure.

Roadmap requirements describe the future UAM concept of operations (e.g., air taxi) that evolves over time from today's similar operations. At the time of this writing, the element of time evolution is defined by UAM Maturity Level (UML) varying from UML-2 to UML-4. UML-2 represents initial UAM operations, while UML-4 envisions mature UAM operations. The characteristics of evolving requirements from UML-2 to higher UMLs is defined by ***derivedReq*** relationship. For instance, onboard pilot requirements in UML-2 evolve into remote pilot requirements in UML-4 implying UML-4 requirements are derived from its corresponding lower UMLs. One roadmap requirement can be a derived requirement from multiple requirements, resulting in many-to-many ***derivedReq*** relationship. Finally, roadmap requirements can be ***allocated*** to system actors who are responsible to execute those requirements. One requirement can have multiple responsible actors, and, likewise, one actor is responsible for multiple requirements, representing many-to-many ***allocated*** relationship.

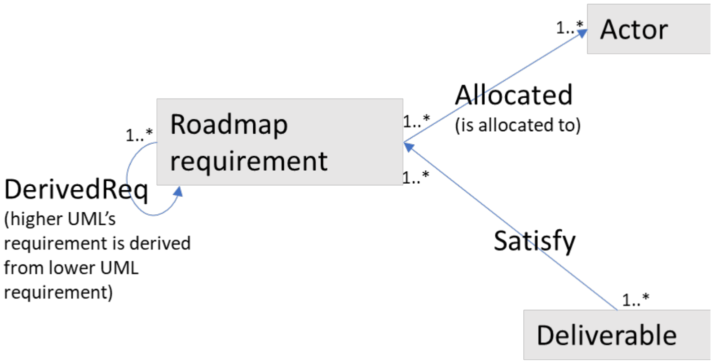


Fig. 2 Knowledge graph ontology.

B. Requirement Maturity Level (RML) Definition

The Requirement Maturity Level (RML) of a research requirement in the Roadmap is a measure of the level of validation of that requirement, based on the set of deliverables that trace to it and their TRL values. The process for assigning a TRL to a deliverable is described in the previous section.

There is a requirement $\mathcal{R}$, and the set of deliverables that are traced to $\mathcal{R}$. The process for tracing a deliverable to a requirement is given in the previous section. Let $i = 1,2 \dots 9$ represent the nine TRLs, and let $n_1, n_2, \dots, n_9$ be the number of deliverables of TRL i that are traced to $\mathcal{R}$. Let $m(\mathcal{R}) = m = \max_{n_i > 0} i$.

It will be assumed that $n_i > 0$ for all $i \leq m$, which is to say there are no deliverables at a high TRL that trace to $\mathcal{R}$ without at least one deliverable at every lower TRL. For example, there is flight tested a requirement that has fast-time low- and high-fidelity simulation behind it.

The RML of $\mathcal{R}$ will be modeled as being a real number x in the range $[0, 10)$, and encode information separately into its integer and fractional components. Let $\mathrm{int}(x) = \lfloor x \rfloor$ and let $\mathrm{frac}(x) = x - \lfloor x \rfloor$. Intuitively, the integer part of the RML should reflect the highest TRL of all deliverables, and the fractional part should indicate how mature that requirement is for the given TRL. That is to say, the fractional part captures the behavior that a requirement with multiple underlying studies that lead to a single flight test is regarded as more mature than a requirement with minimal underlying studies that lead to a flight test.

A family of RML measures is defined to have the following properties:

- A requirement with no associated research deliverables will have an RML value of 0.
- A requirement $\mathcal{R}$ with $m(\mathcal{R}) = m$ will have an RML value in the range $[m, m + 1)$. That is, the integer part of the RML will match the highest TRL of any deliverable traced to $\mathcal{R}$.

---

[4] TRL provides a scale against which to measure the maturity of a technology. TRLs range from 1, basic technology research, to 9, systems test, launch, and operations. (Ref. [2])

- The fractional part of the RML will be a function of the number of all deliverables such that:
    - More deliverables mean higher RML
    - Deliverables at lower TRLs count less towards the RML than deliverables at higher TRLs

A general formula for the proposed family of measures, satisfying the properties above, is given below:

$$RML(\mathcal{R}) = m + \sum_{i=1}^{m} a_i(m) \cdot f(n_i)$$

Where $a_i(m)$ is a scaling factor depending on m, and $f(n_i)$ represents the contribution to the fractional part of the RML by the deliverables with a TRL of i. The following the behavior is modeled: (a) more research means more maturity, and (b) the first few studies provide greatest impact to maturity while later studies at the same TRL contribute diminishing amounts. This is captured by the properties that $f_m(x) \in [0,1]$, is monotonically increasing, and $f''_m(x) < 0$.

The term $a_i(m)$ is an allocation of the relative value of lower-TRL research to a requirement with $m(\mathcal{R}) = m$. It can be chosen arbitrarily if it is monotonically increasing with i and $\sum_1^m a_i(m) = 1$. The dependency on m reflects the fact that the relative contribution of lower-level research is greater for less mature requirements. For example, a fast-time feasibility study (TRL=2) that traces to $\mathcal{R}$ will have greater relative contribution if $m(\mathcal{R}) = 2$ than if $m(\mathcal{R}) = 5$.

An application of the formula is provided with the following specific properties:

- Research at TRL i contributes about twice as much as research at TRL $i - 1$
- The first deliverable at any TRL does not increase the fractional part of the RML. This captures the behavior that getting the first deliverable at a higher TRL will increase the integer value and not also increase the fractional value.
- After the first deliverable, the value of additional research at the same TRL reduces by a factor of two.

$$a_i(m) = \frac{1}{2^{m-i}} \frac{2^{m-1}}{2^m - 1} = \frac{2^{i-1}}{2^m - 1}$$

$$f(n_i) = 1 - \frac{1}{2^{n_i - 1}}$$

$$RML(\mathcal{R}) = m + \sum_{i=1}^{m} \frac{2^{i-1}}{2^m - 1} \cdot \left(1 - \frac{1}{2^{n_i - 1}}\right)$$

To get a sense of how the equation works, the table below lists out the allocations $a_i(m)$. The columns are the values of the a_i for $i = 1, 2, \dots m$ for each value of m. Note that the columns sum to one.

i,m	1	2	3	4	5	6	7	8	9
1	1.0	0.33	0.14	0.07	0.03	0.02	0.01	0.00	0.00
2		0.67	0.29	0.13	0.06	0.03	0.02	0.01	0.00
3			0.57	0.27	0.13	0.06	0.03	0.02	0.01
4				0.53	0.26	0.13	0.06	0.03	0.02
5					0.52	0.25	0.13	0.06	0.03
6						0.51	0.25	0.13	0.06
7							0.50	0.25	0.13
8								0.50	0.25
9									0.5

The following table demonstrates the behavior of this RML measure through a scenario of research progression, characterized by the build-up of deliverables. The table below should be read from left to right, where moving from column k to column $k + 1$ represents the addition of new deliverables to the research. Note that there is a specific column for every time the first high-TRL deliverable is added, showing the case where the integer value of RML increases by one.

n_1	1	4	4	5	5	7	7	8	8	10
n_2			1	3	3	4	4	5	5	7
n_3					1	1	2	2	3	4
n_4								1	2	3
n_5										1
RML	**1**	**1.9**	**2.3**	**2.8**	**3.3**	**3.4**	**3.7**	**4.3**	**4.7**	**5.4**

C. Knowledge Graph Data Science

Knowledge graph was first introduced by Google in 2012 (Ref. [4]) to enable its users to discover new information quickly and relevant to their queries. It was based on graph theory, which is a mathematical concept that classifies elements in terms of vertices (nodes) and edges (relationships) to understand connections and patterns within the information being studied. Knowledge graph is stored in the KbDP as a graph database. Several graph algorithms can be applied for various use cases. For this study, the PageRank Centrality algorithm is used to identify roadmap requirement's importance.

PageRank (PR) (Ref. [5]) is a Centrality algorithm used by Google Search to rank web pages in their search engine results. PageRank measures the importance of webpages by calculating a ratio of incoming and outgoing links, taking into account the respective importance of corresponding linked pages. Intuitively, pages that are well cited from many places around the web are worth looking at (high score). Also, pages that have only one citation from a respected and quality source are also generally worth looking at.

In this study, PageRank scores of requirements based on ***derivedReq*** relationships are calculated. Heuristically, UML-3 requirement is derived from UML-2, therefore it is more critical to ensure its UML-2 requirement is thoroughly validated. Additionally, PageRank scores are also based on ***allocated*** relationships because any requirement with multiple responsible actors should be more critical to be thoroughly validated. A notional example of how these scores are computed can be depicted in Figure 3. Intuitively, requirement node 1 has more incoming links than those of node 2, therefore, the PageRank score of node 1 (representing by its size) is larger than that of node 2. Additionally, node 3 has only one incoming link but the source of this incoming link has a high score, therefore, node 3 will also have a high score. In this example, both nodes 1 and 3 are more important than node 2. Following this example, using this score all roadmap requirements can be ranked for its relative importance or criticality.

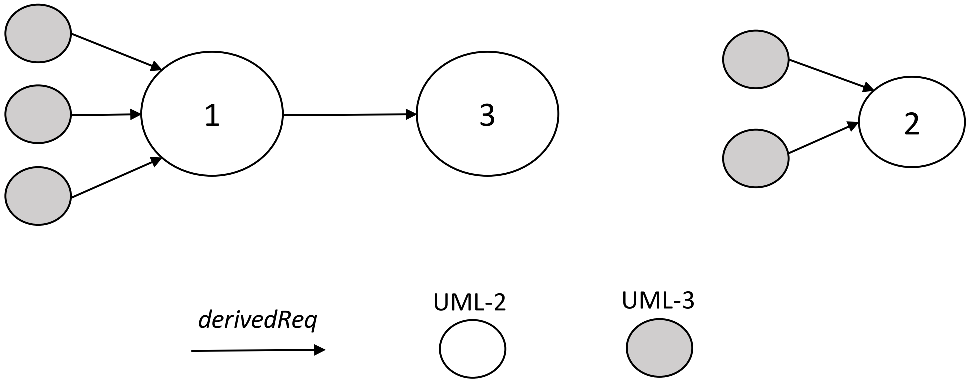


Fig. 3 Notional PageRank Score.

D. Portfolio Analysis

The PageRank scores will identify critical airspace roadmap requirements. Then the RML values of these important requirements will be inspected to determine whether or not there is any to-date research deliverable maturing them. Any important requirement with low RML provides the opportunity for future research portfolio investment.

III. Analysis Results

This section presents analysis results from methodology described in the previous section. Subsection A gives overall descriptive statistics to understand the knowledge graph size. The PageRank scores are analyzed to determine top important requirements in Subsection B. Lastly, Subsection C investigates the top important requirements' RML values and determines the opportunity for future research portfolio investment.

A. Knowledge Graph Size

There are 340 roadmap requirements, where 313 have incoming and outgoing relationships and 27 have no relationship. Further inspection found that these 27 requirements are all "To Be Resolved" where they represent best estimates, a lack of known requirements, assumptions, or constraints, or simply areas where further development is needed. For this study, only 313 requirements are considered.

Similarly, there are 25 deliverables, where 22 have outgoing (***satisfy***) relationships and 3 have no relationship. Further inspection found that these 3 deliverables are not yet completed (at the time of this writing) but listed in the database as planned for future completion. Therefore, only 22 deliverables are used in this study. There are 12 actors, and all actors have incoming (***allocated***) relationships. There are 515 ***allocated*** relationships, 508 ***satisfy*** relationships, and 948 ***derivedReq*** relationships.

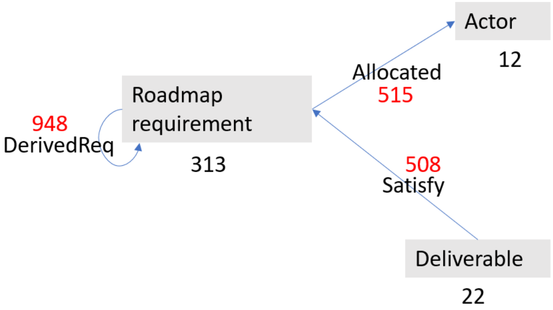


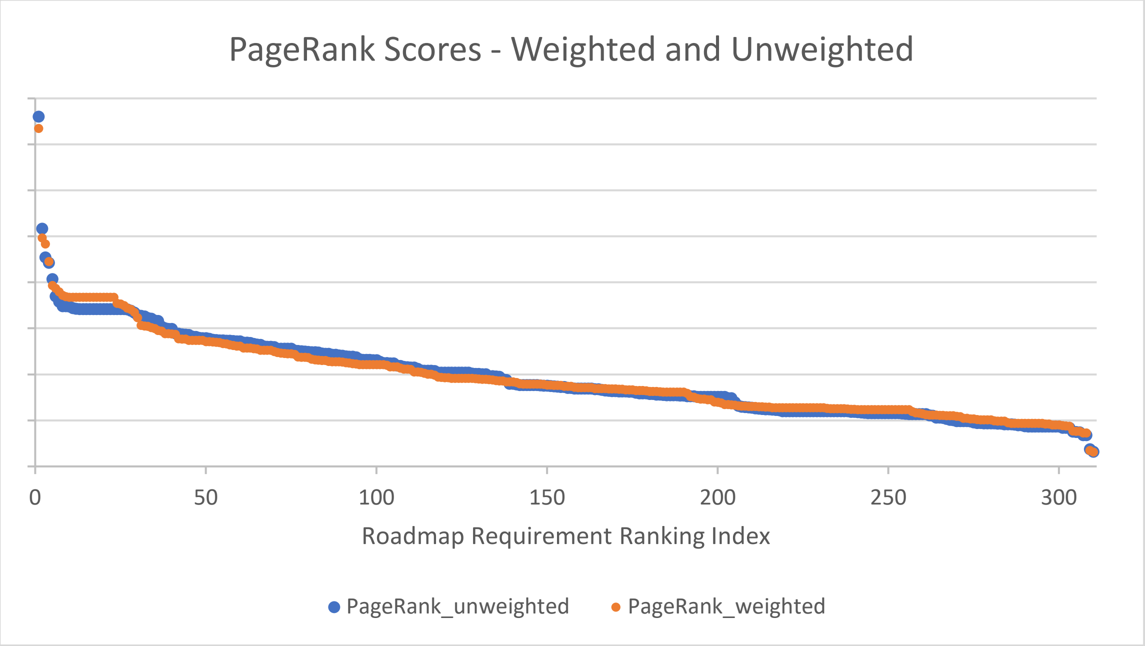
Fig. 4 Knowledge Graph Size.

Generally, each deliverable satisfies on average of about 23 (508/22) requirements. On average, each requirement is validated by 1.6 (508/313) deliverables. Each requirement has on average 1.6 (515/313) responsible actors. On average, each actor is responsible for about 43 (515/12) requirements. Finally, each requirement is derived from an average of 3 (948/313) other requirements.

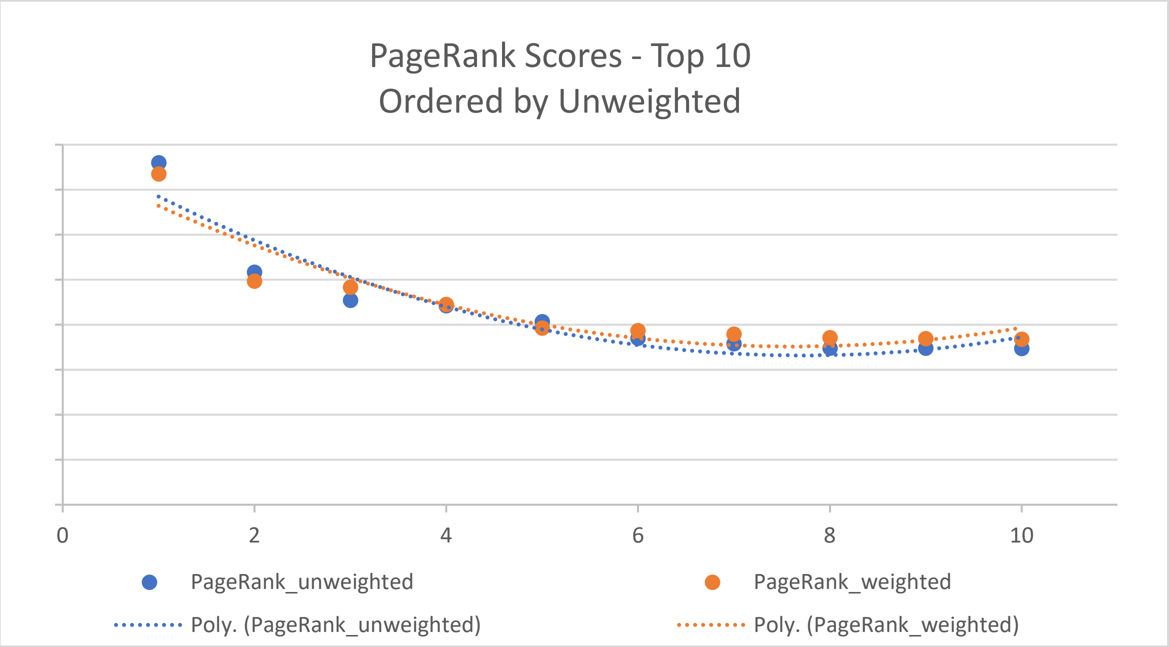
B. Important Requirement Identification

The PageRank Centrality algorithm is executed on the sample of 313 roadmap requirements, 12 actors, 948 ***derivedReq***, and 515 ***allocated*** edges. Because there are twice as many ***derivedReq*** edges as ***allocated***, the ***derivedReq*** edges may trump the ***allocated*** edges, likely producing biased results. For this reason, normalization of these edges is implemented through the weighting of the ***allocated*** edge by a factor of 1.84 (948/515). Nevertheless, this study continues to explore the sensitivity of weighted and unweighted cases.

Figure 5a shows the sorted PageRank scores for all 313 requirements for weighted and unweighted cases. The scores drop sharply initially and then gradually and asymptotically approaching a minimum non-zero PageRank default value set by the algorithm. This behavior implies that there is a finite set of important requirements. Figure 5b zooms into the top 10 scores with the overlaying polynomial trendlines, indicating the top 5 requirements should be considered as the most important.



(a) Sorted PageRank Scores for all 313 requirements.



(b) Sorted PageRank Scores for top 10 scores.

Fig. 5 Sorted PageRank Scores.

Table 1 shows the top 5 requirements for both weighted and unweighted cases. Both top requirement sets are the same except the ranking placement, i.e., the 3rd and 4th places are reversed. Further inspection reveals that these top five are highly related to number of allocated actors (see Figure 6). In other words, any requirement with abnormally high number of allocated actors (in this case, 6 or 7 actors) will be important. Breaking number of actors' tie to determine additional important requirements will be determined by the ***derivedReq*** edges.

Figure 7 displays the top five requirement's tree diagrams, in which the tree diagram presents various types of nodes and links. Requirement 194 UML-3.AD has 7 allocated actors and 13 other roadmap requirements that were derived from it, resulting in being the most important requirement. In other words, if this requirement has no research to validate it, then its 13 derived requirements could become invalid.

Requirement 150 UML-3.AD has 4 allocated actors and 12 other requirements that were derived from it, placing it as the 2nd most important requirement. Requirement 89 UML-4.SS has 6 allocated actors and 2 other roadmap requirements depending on it, placing it as the 3rd or 4th most important. Requirement 36 UML-3.AM has 4 allocated actors and 12 other dependable requirements, resulting in being the 4th or 3rd most important. Lastly, requirement 336 UML-4.CS has 4 allocated actors and 9 other dependable requirements, placing it as the 5th most important.

The requirement descriptions (given in Table 1) reveal that requirements relate to the contingency, off-nominal, and emergency operations are identified as important (the 1st, 2nd, and 5th places). Additionally, requirements relate to the separation responsibility and the information sharing are important (the 3rd and 4th places).

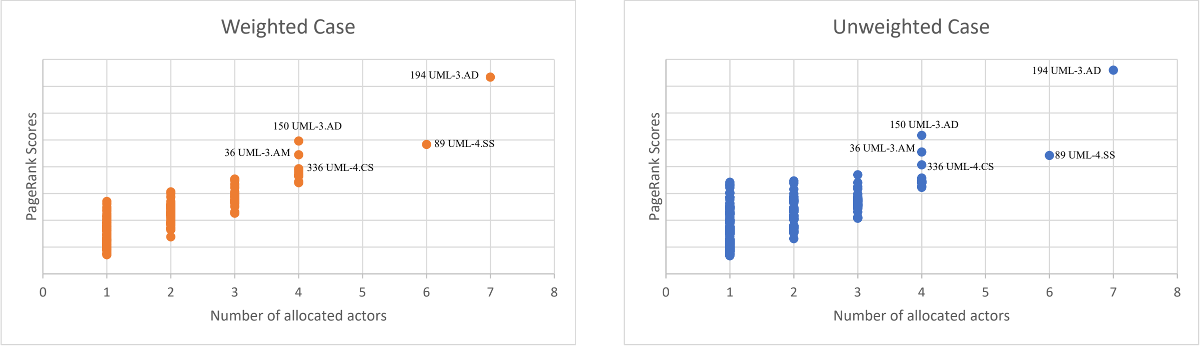


Fig. 6 Direct correlation between the PageRank scores and the number of allocated actors.

Table 1 Summary of the top 5 requirements identified by the PageRank scores.

Requirement ID	Weighted – Placement	Unweighted – Placement	Full Requirement Text
194 UML-3.AD	1st	1st	The FAA will provide a means to authorize contingency procedures that include the UAM Operator, Vertiport Operator, Fleet Manager, Vertiport Manager, PIC, or ATC.
150 UML-3.AD	2nd	2nd	The UAM Operator shall establish a plan to resolve off-nominal conditions with the PIC, Fleet Manager, or Vertiport Manager, without ATC involvement.
89 UML-4.SS	**3rd**	**4th**	The FAA will approve flight rules that allow ATC to delegate separation responsibility to the PIC, Fleet Manager, or Vertiport Manager, or PSU under appropriate conditions.
36 UML-3.AM	**4th**	**3rd**	The PSU, UAM Operator, Vertiport Operator shall discover the information needed to build a Common Operating Picture.
336 UML-4.CS	5th	5th	The RPIC should communicate with the Fleet Manager, Vertiport Manager, or ATC to resolve contingency and emergency operations.

C. Research Portfolio Assessment

Figure 7 displays not only the top five requirement's tree diagrams identified in the previous section, but also its associated RML value. There is no deliverable node associated with the 194 UML-3.AD and 336 UML-4.CS resulting in its RML of zero. In fact, at the time of this writing there is no ATM-X UAM research addressing the contingency and emergency operations in the current research portfolio. The 150 UML-3.AD has one deliverable, called "AIS deliverables", that will satisfy this requirement; however, this deliverable has not yet completed at the time of this writing, and therefore the RML is zero. Once the deliverable is complete, this requirement will have a non-zero RML.

Both the 89 UML-4.SS and 36 UML-3.AM requirements have 1 and 8 associated deliverables, respectively. The more deliverables satisfying a requirement, the higher its RML is, as seen in the figure that the RML of 2 for the 89 UML-4.SS and 3.7 for 36 UML-3.AM. Potential future research could focus on the contingency and emergency operations to address the two important requirements without any associated deliverable as identified in this study.

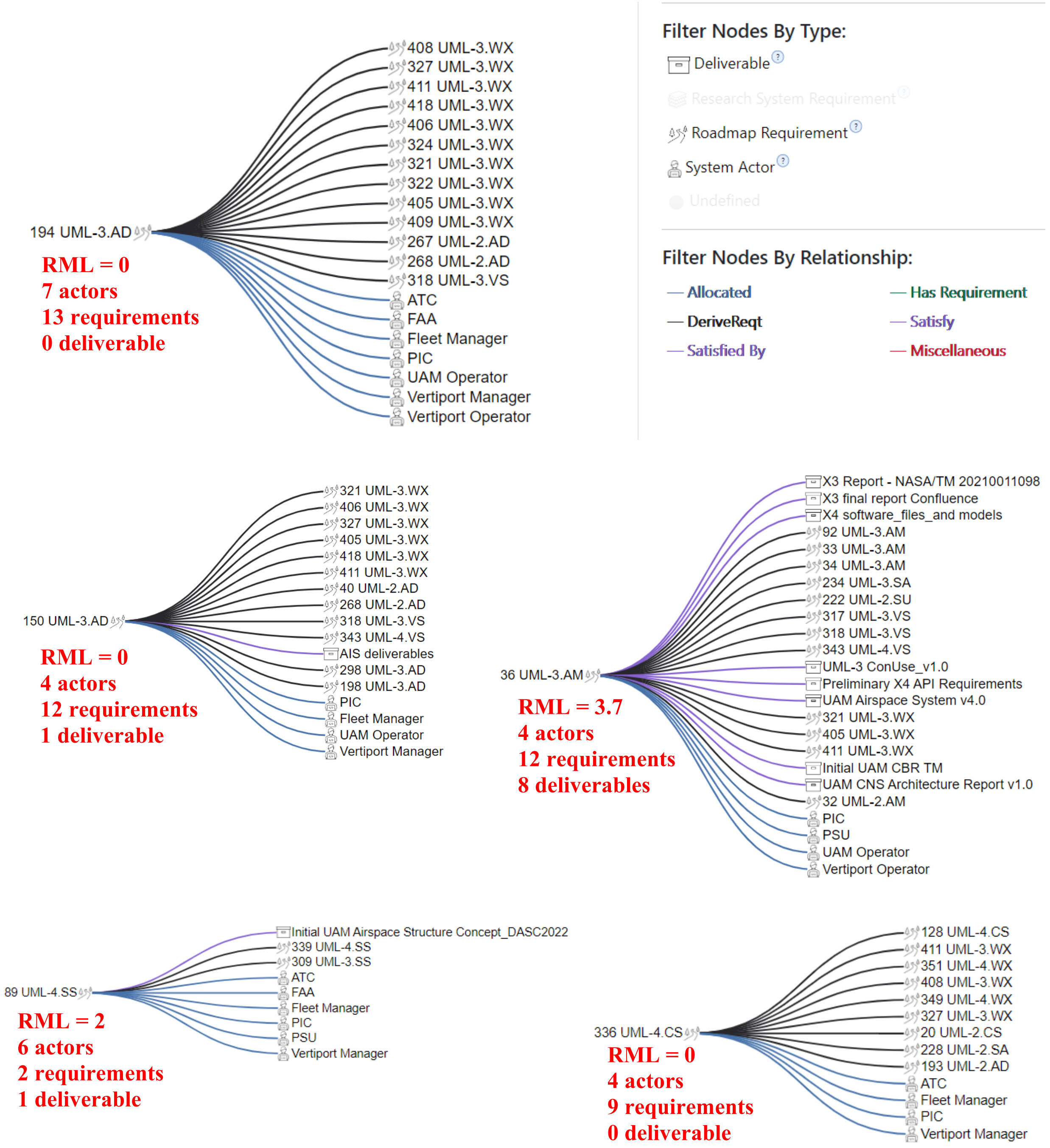


Fig. 7 Top five important requirements based on PageRank with its associated RML value.

IV. Conclusion and Future Work

The advanced data analytics in knowledge graph was applied to the existing UAM knowledge database and was successfully used to assess the gap in the current research portfolio for the project. The methodology presented in this study is an initial step of knowledge graph data science application to identify research gap for the project. Much work is still needed to be done. This database is rich with complex relationships, having more node and relationship types than what were presented in this study. At the time of this writing, the current knowledge database has 937 nodes and 2,856 links, about 1.6 times the size in this study. Potential next step is to further mature this methodology for other node and relationship types. If successful, this methodology will greatly assist in research gap analysis, which in turn significantly help project managers in replanning future research portfolio not only for the UAM operations but also for the AAM R&D efforts.

References

[1] Levitt, I., Phojanamongkolkij, N., Horn, A., and Witzberger, K., "UAM Airspace Research Roadmap, Revision 2.0," NASA TM- 2023-0002647, 2023.

[2] National Aeronautics and Space Administration, *NASA Systems Engineering Handbook*, NASA/SP-2016-6105 Rev 2, 2016.

[3] National Aeronautics and Space Administration, https://esto.nasa.gov/trl/ (Accessed: 18 October 2023).

[4] Singhal, A., "Introducing the Knowledge Graph: thinks, not strings," https://blog.google/products/search/introducing-knowledge-graph-things-not/ (Accessed: 18 October 2023).

[5] https://en.wikipedia.org/wiki/PageRank (Accessed: 18 October 2023).